Decoupled image-based visual servoing for cameras obeying the unified projection model

Omar Tahri, Youcef Mezouar, François Chaumette, Peter Corke

Decoupled Image-Based Visual Servoing for Cameras Obeying the Unified Projection Model

Omar Tahri, Youcef Mezouar, François Chaumette, *Senior Member, IEEE*, and Peter Corke

***Abstract*—This paper proposes a generic decoupled image-based control scheme for cameras obeying the unified projection model. The scheme is based on the spherical projection model. Invariants to rotational motion are computed from this projection and used to control the translational degrees of freedom (DOFs). Importantly, we form invariants that decrease the sensitivity of the interaction matrix to object-depth variation. Finally, the proposed results are validated with experiments using a classical perspective camera as well as a fisheye camera mounted on a 6-DOF robotic platform.**

***Index Terms*—Decoupling, invariants, omnidirectional cameras, visual servoing.**

I. Introduction

IN IMAGE-BASED visual servoing (IBVS), the choice of the set of visual features to be used in the control scheme is still an open question. Image features that can be, and have been, used include the coordinates of interest points, the parameters of lines or conics, the moments of planar patches, etc. We wish to choose features that lead to control behavior that is optimal with respect to the image, e.g., keeping points within the field of view, as well as with respect to 3-D camera motion, e.g., minimal distance moved and avoidance of robot singularities.

The choice of visual features has a strong influence on the performance of the control system and on the ability to analyze the system's dynamics. A common approach using simple features is image-point-based visual servoing. While theoretically suitable only for "small" displacements, in practice, it is quite robust but less than optimal in terms of 3-D motion [3].

One way to improve performance is to sample the initial errors to ensure that the error at each iteration remains small in order to overcome the problems that are mentioned previously. This combines a path-planning process together with the servoing one [4], [5], [9], [21], [23]. A second way involves more modeling of the nonlinearities in the relationship between the image and workspace. Lapreste and Mezouar [15] present a method to estimate the control matrix in visual servoing using a second-order approximation of the projection function based on a Hessian approximation. The main drawback is that this method introduces a number of supplementary parameters. In order to avoid the Hessian computation, an efficient method combining the desired and the current values of the interaction matrix is proposed in [18]. The latter method has been improved in [28] by taking into account the tensor change of frames.

We are interested in approaches that consider performance measures to choose visual features with good decoupling and linearizing properties. In fact, the choice of features directly influences the closed-loop dynamics in task-space. Several works have been realized in image-based visual servoing following the same general objective. Features including the distance between two points in the image plane and the orientation of the line connecting those two points were proposed in [8]. The relative area of two projected surfaces was proposed in [31] as a feature. A vanishing point and the horizon line were selected in [22], which ensures good decoupling between translational and rotational degrees of freedom (DOFs). Vanishing points have also been used in [16] for a dedicated object (e.g., a rectangle), once again to obtain some decoupling properties. For the same object, six visual features have been designed in [6] to control the six DOFs of a robot arm, following a partitioned approach. The coordinates of points are expressed in a cylindrical coordinate system instead of the classical Cartesian one in [14] to improve the robot trajectory. In [13], the three coordinates of the centroid of an object in a virtual image obtained through a spherical projection have been selected to control three DOFs of an underactuated system. Mahony *et al.* [17] deal with the selection of the optimal feature to control the camera motion with respect to the depth axis. Fomena and Chaumette [30] propose a decoupled visual servoing from spheres using a spherical projection model. Despite many reported results over the past few years, the choice of the set of visual features to be used in the control scheme is still an open question. Performance criteria to choose feature sets include stability and validity for different kinds of sensors and environments.

Image moments have been widely studied in the computer-vision community, especially for pattern-recognition applications. Indeed, invariance to some transformations such as scale, 2-D translation, and/or 2-D rotation can be obtained by appropriate combinations of moments. Moment invariants have been well studied for pattern recognition; for instance, see [12], [19], and [25]. This invariance property is also of particular interest in visual servoing. By selecting an appropriate combination

Manuscript received August 28, 2009; revised March 19, 2010; accepted May 23, 2010. Date of publication July 1, 2010; date of current version August 10, 2010. This paper was recommended for publication by Associate Editor D. Kragic and Editor G. Oriolo upon evaluation of the reviewers' comments.

O. Tahri is with the Institute for Systems and Robotics, Polo II 3030-290 Coimbra, Portugal (e-mail: omartahri@isr.uc.pt).

Y. Mezouar is with the LASMEA, Université Blaise Pascal, Campus des Cezeaux, 63177 Clermont-Ferrand, France (e-mail: mezouar@lasmea.univ-bpclermont.fr).

F. Chaumette is with the INRIA Rennes, Bretagne Atlantique, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France (e-mail: francois.chaumette@irisa.fr).

P. Corke is with the School of Engineering Systems, Queensland University of Technology, 4001 Brisbane, Qld., Australia (e-mail: peter.corke@qut.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2010.2051593

of moments, it becomes possible to create partitioned systems with good decoupling and linearizing properties [25], [26]. For instance, by using such features, the interaction-matrix block corresponding to the translational velocity can be a block diagonal with no depth dependence. However, this approach is limited to planar objects and conventional perspective cameras. A new decoupled image-based control scheme using the projection onto a unit sphere has been proposed in [27], which is based on polynomials invariant to rotational motion computed from a set of image points. More recently, a decoupled image-based control scheme based on the surface of triangle projection onto a sphere has been proposed in [29]. This paper synthesizes our contributions while developing the theoretical and experimental results. In particular, the computation of the interaction matrix related to the projection surface of triangles is detailed, and its invariance to rotations is formally shown. This paper also provides a new and complete set of real experiments as well as new simulation results. The proposed control schemes are not only compared between them but are also compared with an image-based control scheme using points coordinates as visual features.

As mentioned above, the features we propose are computed from the projection onto the unit sphere. This means that the proposed method can work not only with classical perspective cameras but can also be applied to wide-angle cameras obeying the unified model [2], [10]. Wide-angle cameras include catadioptric systems that combine mirrors and conventional cameras to create omnidirectional cameras providing 360° panoramic views of a scene or dioptric fisheye lenses [1]. It is highly desirable that such imaging systems have a single viewpoint [1], [24], i.e., there exists a single center of projection so that every pixel in the sensed images measures the irradiance of the light passing through the same viewpoint in one particular direction. The reason why a single viewpoint is so desirable is that it permits the extension of several results obtained for conventional cameras [11]. In this paper, we also take advantage of the properties of such sensor system to develop control laws that are valid for conventional, catadioptric, and fisheye cameras.

In the next section, we recall the unified camera model and the control law. Moment definitions and the interaction matrices computation are also presented. In Section III, theoretical details about feature selection are discussed, and a new vector of features to control the 6-DOF camera is proposed. Finally, in Section IV, experimental results obtained using a conventional camera and a fisheye camera mounted on a 6-DOF robot are presented to validate our approach.

II. Modeling

A. Camera Model

Central imaging systems can be modeled using two consecutive projections: first spherical and then perspective. This geometric formulation, which is called the *unified model*, was proposed by Geyer and Daniilidis [10]. Let us consider a virtual unitary sphere centered on C_m and the perspective camera centered on C_p (see Fig. 1). The frames attached to the sphere and the perspective camera are related by a simple translation of $-\xi$ along the Z-axis. Let $\mathcal{X}$ be a 3-D point with coordinates

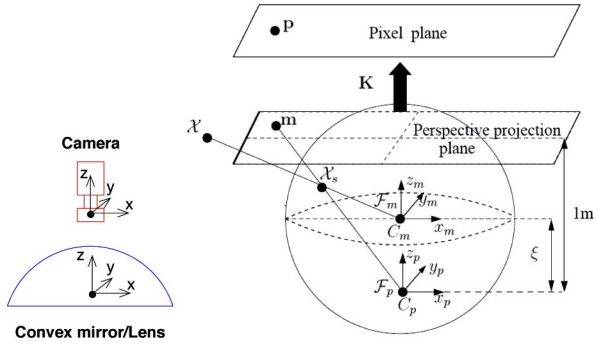


Fig. 1. (Left) Catadioptric camera and mirror geometry. (Right) Unified image formation.

$\mathcal{X} = (X, Y, Z)$ in $\mathcal{F}_m$. The world point $\mathcal{X}$ is projected to

$$\mathbf{m} = (\, x, \quad y, \quad 1\,) = \left(\frac{X}{Z + \xi \|\mathcal{X}\|}, \quad \frac{Y}{Z + \xi \|\mathcal{X}\|}, \quad 1 \right) \tag{1}$$

and then mapped to the homogeneous image-plane coordinate $\mathbf{p} = \mathbf{K}\mathbf{m}$, where $\mathbf{K}$ is a 3×3 matrix of camera and mirror intrinsic parameters. The matrix $\mathbf{K}$ and the parameter ξ can be obtained after calibration using, for example, the methods proposed in [20]. In the sequel, the imaging system is assumed to be calibrated. In this case, the inverse projection onto the unit sphere can be obtained by

$$\mathcal{X}_{\mathbf{s}} = \lambda \left(x, \quad y, \quad 1 - \frac{\xi}{\lambda} \right) \tag{2}$$

where

$$\lambda = \frac{\xi + \sqrt{1 + (1 - \xi^2)(x^2 + y^2)}}{1 + x^2 + y^2}.$$

Note that the conventional perspective camera is nothing but a particular case of this model where $\xi = 0$. The projection onto the unit sphere from the image plane is possible for all sensors obeying the unified model.

B. Image-Based Visual Servoing

We define the vector of image features $\mathbf{s}$ and recall that its time variation

$$\dot{\mathbf{s}} = \mathbf{L}_{\mathbf{s}} \mathbf{V} \tag{3}$$

is linear with respect to the relative camera–object kinematics screw $\mathbf{V} = (\mathbf{v}, \boldsymbol{\omega})$, and $\mathbf{L}_{\mathbf{s}}$ is the interaction matrix related to $\mathbf{s}$. The control scheme is usually designed to reach an exponential decoupled convergence of the visual features to their desired value $\mathbf{s}^*$ [7]. If we consider an eye-in-hand system observing a static object, the control law is defined as follows:

$$\mathbf{V}_{\mathbf{c}} = -\lambda \widehat{\mathbf{L}_{\mathbf{s}}}^{+} (\mathbf{s} - \mathbf{s}^*) \tag{4}$$

where $\widehat{\mathbf{L}_{\mathbf{s}}}$ is a model or an approximation of $\mathbf{L}_{\mathbf{s}}$, $\widehat{\mathbf{L}_{\mathbf{s}}}^{+}$ is the pseudoinverse of $\widehat{\mathbf{L}_{\mathbf{s}}}$, λ is a positive gain, and $\mathbf{V}_{\mathbf{c}}$ is the camera velocity sent to the low-level robot controller. Equation (4) is a linear approximation of the nonlinear mapping between 3-D and image space and, therefore, valid for small displacements.

However, for large displacements, the approximation is not valid and can lead to suboptimal robot trajectories.

An important issue is therefore to determine those visual features that will allow the system dynamics to be linear over large displacements. Furthermore, using (4), local minima can be reached when the number of features is not minimal. Therefore, one would like to choose a minimal representation (i.e., the number of features is equal to the number of DOFs) but without singularities and robust with respect to noise in the image.

C. Invariants to Rotational Motions From the Projection Onto the Surface of Unit Sphere

The shape of a planar object does not change under rotational motions. After a rotational motion of the sensor frame, it can easily be shown that the projected shape undergoes the same rotational motion as the coordinates of the object 3-D points. This means that the invariants to rotation in 3-D space are also invariant if the considered points are projected onto the unit sphere. The decoupled features we propose are based on this invariance property. It will be used to select features invariant to rotations in order to control the three translational DOFs. In this way, the following polynomial that is invariant to rotations has been proposed in [27] to control the translational DOFs:

$$I_1 = m_{200}m_{020} - m_{200}m_{002} + m_{110}^2 + m_{101}^2 - m_{020}m_{002} + m_{011}^2 \tag{5}$$

where $m_{i,j,k}$ is the 3-D moment of order $i+j+k$ computed from a discrete set of points defined by the following classical equation:

$$m_{i,j,k} = \sum_{h=1}^{N} x_h^i \; y_h^j \; z_h^k \tag{6}$$

where $(x_h,\, y_h,\, z_h)$ is the coordinates of the hth point, and N is the number of points. In our case, these coordinates are nothing but the coordinates of a point projected onto the unit sphere. In this paper, another kind of invariant is derived from the projection onto the unit sphere. More precisely, the surface that can be computed from the projection of three noncollinear points onto the unit sphere will also be used. In this case, two kinds of surfaces that are invariants to rotations can be defined: the surface defined by the triangle projected onto the unit sphere (which is defined by three circular arcs corresponding to the projection of the triangle's edges onto the unit sphere) and the surface Δ of the triangle formed by projection onto the sphere (see Fig. 2). The latter surface is computed by the well-known formula for triangle surface

$$\Delta = \frac{1}{2}\|(\mathcal{X}_{s_2} - \mathcal{X}_{s_1}) \times (\mathcal{X}_{s_3} - \mathcal{X}_{s_1})\| \tag{7}$$

where $\mathcal{X}_{s_1} = (x_{s_1}, y_{s_1}, z_{s_1})$, $\mathcal{X}_{s_2} = (x_{s_2}, y_{s_2}, z_{s_2})$, and $\mathcal{X}_{s_3} = (x_{s_3},\, y_{s_3},\, z_{s_3})$ are the coordinates of the triangle's vertices projected onto the unit sphere.

In the following, it is this surface Δ that will be used. We will show that after an adequate transformation, new features can be obtained from Δ, as well as from I_1 given by (5), such that the corresponding interaction matrices are almost constant with respect to variation of object depth. A comparison of the use of the two features will be made.

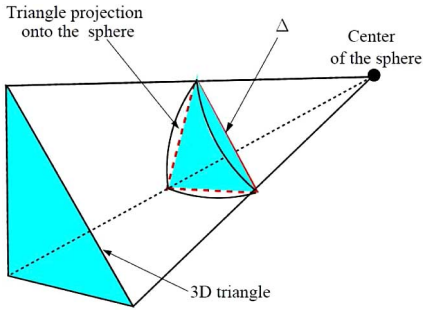


Fig. 2. Triangle projection onto the unit sphere.

D. Interaction Matrix

In the case of moments computed from a discrete set of points, the derivative of (6) with respect to time is given by

$$\dot{m}_{i,j,k} = \sum_{h=1}^{N} \big(i\; x_{s_h}^{i-1}\; y_{s_h}^{j}\; z_{s_h}^{k}\, \dot{x}_{s_h} + j\; x_{s_h}^{i}\; y_{s_h}^{j-1}\; z_{s_h}^{k}\, \dot{y}_{s_h} + k\; x_{s_h}^{i}\; y_{s_h}^{j}\; z_{s_h}^{k-1} \dot{z}_{s_h} \big). \tag{8}$$

The interaction matrix $\mathbf{L}_{\mathcal{X}_s}$ for a point on the unit sphere is well known [13], [25], [30] and is given by

$$\mathbf{L}_{\mathcal{X}_s} = \Big[-\frac{1}{r}\mathbf{I}_3 + \frac{1}{r}\mathcal{X}_s\mathcal{X}_s^{\top} \quad [\mathcal{X}_s]_{\times} \Big] \tag{9}$$

where r is the distance of the 3-D point to the sphere center. For any set of points (i.e., coplanar or noncoplanar), we can combine that interaction matrix with that related to $\mathbf{L}_{\mathbf{m}_{\mathbf{i,j,k}}}$, from (8), to obtain

$$\mathbf{L}_{m_{i,j,k}} = [\, m_{vx} \quad m_{vy} \quad m_{vz} \quad m_{wx} \quad m_{wy} \quad m_{wz} \,] \tag{10}$$

where

$$\begin{cases} m_{vx} = \displaystyle\sum_{h=1}^{N} \frac{-i x_{s_h}^{i-1}\; y_{s_h}^{j}\; z_{s_h}^{k} + \beta_d x_{s_h}^{i+1} y_{s_h}^{j}\; z_{s_h}^{k}}{r_h} \\ m_{vy} = \displaystyle\sum_{h=1}^{N} \frac{-j x_{s_h}^{i}\; y_{s_h}^{j-1}\; z_{s_h}^{k} + \beta_d x_{s_h}^{i}\; y_{s_h}^{j+1} z_{s_h}^{k}}{r_h} \\ m_{vz} = \displaystyle\sum_{h=1}^{N} \frac{-k x_{s_h}^{i}\; y_{s_h}^{j}\; z_{s_h}^{k-1} + \beta_d x_{s_h}^{i}\; y_{s_h}^{j}\; z_{s_h}^{k+1}}{r_h} \\ m_{wx} = j m_{i,j-1,k+1} - k m_{i,j+1,k-1} \\ m_{wy} = k m_{i+1,j,k-1} - i m_{i-1,j,k+1} \\ m_{wz} = i m_{i-1,j+1,k} - j m_{i+1,j-1,k} \end{cases}$$

and $\beta_d = i+j+k$. In the particular case of a coplanar set of points, the interaction matrix related to $m_{i,j,k}$ can be determined

[25] as follows:

$$\begin{cases} m_{vx} = A(\beta_d m_{i+2,j,k} - im_{i,j,k}) \\ \qquad + B(\beta_d m_{i+1,j+1,k} - im_{i-1,j+1,k}) \\ \qquad + C(\beta_d m_{i+1,j,k+1} - im_{i-1,j,k+1}) \\ m_{vy} = A(\beta_d m_{i+1,j+1,k} - jm_{i+1,j-1,k}) \\ \qquad + B(\beta_d m_{i,j+2,k} - jm_{i,j,k}) \\ \qquad + C(\beta_d m_{i,j+1,k+1} - jm_{i,j-1,k+1}) \\ m_{vz} = A(\beta_d m_{i+1,j,k+1} - km_{i+1,j,k-1}) \\ \qquad + B(\beta_d m_{i,j+1,k+1} - km_{i,j+1,k-1}) \\ \qquad + C(\beta_d m_{i,j,k+2} - km_{i,j,k}) \\ m_{wx} = jm_{i,j-1,k+1} - km_{i,j+1,k-1} \\ m_{wy} = km_{i+1,j,k-1} - im_{i-1,j,k+1} \\ m_{wz} = im_{i-1,j+1,k} - jm_{i+1,j-1,k} \end{cases} \tag{11}$$

where $\alpha = (A,\ B,\ C)$ are the parameters defining the object plane in the camera frame

$$\frac{1}{r} = \boldsymbol{\alpha}^\top \mathcal{X}_s = Ax_s + By_s + Cz_s. \tag{12}$$

The interaction matrix related to Δ can be obtained in a similar way. Let $\mathbf{L}_{\boldsymbol{\Delta}}$ be the 1×6 interaction matrix related to Δ and is given by

$$\mathbf{L}_{\boldsymbol{\Delta}} = [\mathbf{L}_{\boldsymbol{\Delta}_v}\ \ \mathbf{L}_{\boldsymbol{\Delta}\omega}] \tag{13}$$

where $\mathbf{L}_{\boldsymbol{\Delta}_v}$ and $\mathbf{L}_{\boldsymbol{\Delta}\omega}$ are, respectively, two 1×3 matrices that link the time variation of Δ to the translational and the rotational velocities.

Lemma 1: Δ is invariant to rotations, and

$$\mathbf{L}_{\boldsymbol{\Delta}\omega} = [\mathbf{0},\ \mathbf{0},\ \mathbf{0}].$$

Proof: Let $\mathbf{L}_{\mathcal{X}_i}$ be the interaction matrix related to the point $\mathcal{X}_{s_i}$, $\mathbf{L}_{\mathcal{X}_{ij}} = \mathbf{L}_{\mathcal{X}_i} - \mathbf{L}_{\mathcal{X}_j}$ be the interaction-matrix difference, and $\mathcal{X}_{ij} = \mathcal{X}_{s_i} - \mathcal{X}_{s_j}$ be the coordinate-vector difference. The surface Δ can be written as

$$\Delta = \frac{\sqrt{\mathcal{X}_{31}^\top [\mathcal{X}_{21}]_\times^\top [\mathcal{X}_{21}]_\times \mathcal{X}_{31}}}{2} = \frac{\sqrt{\mathcal{X}_{21}^\top [\mathcal{X}_{31}]_\times^\top [\mathcal{X}_{31}]_\times \mathcal{X}_{21}}}{2}. \tag{14}$$

Taking the time derivative of (14), we obtain

$$\mathbf{L}_\Delta = \frac{\mathcal{X}_{31}^\top [\mathcal{X}_{21}]_\times^\top [\mathcal{X}_{21}]_\times \mathbf{L}_{\mathcal{X}_{31}} - \mathcal{X}_{31}^\top [\mathcal{X}_{21}]_\times^\top [\mathcal{X}_{31}]_\times \mathbf{L}_{\mathcal{X}_{21}}}{4\Delta}. \tag{15}$$

Combining (15) with (9), it follows that

$$\mathbf{L}_{\Delta_\omega} = \frac{\mathcal{X}_{21}^\top [\mathcal{X}_{31}]_\times^\top ([\mathcal{X}_{21}]_\times [\mathcal{X}_{\mathbf{31}}]_\times - [\mathcal{X}_{31}]_\times [\mathcal{X}_{\mathbf{21}}]_\times)}{4\Delta}. \tag{16}$$

Additionally, it can easily be shown that

$$[\mathcal{X}_{21}]_\times [\mathcal{X}_{\mathbf{31}}]_\times - [\mathcal{X}_{31}]_\times [\mathcal{X}_{\mathbf{21}}]_\times = [\mathcal{X}_{21} \times \mathcal{X}_{31}]_\times. \tag{17}$$

Let us consider $\mathbf{h} = \mathcal{X}_{21} \times \mathcal{X}_{31}$, which allows (16) to be written as

$$\mathbf{L}_{\Delta_\omega} = \frac{\mathbf{h}^\top [\mathbf{h}]_\times}{4\Delta} \tag{18}$$

from which we immediately deduce

$$\mathbf{L}_{\Delta_\omega} = [\mathbf{0}\ \ \mathbf{0}\ \ \mathbf{0}] \tag{19}$$

which confirms the invariance of Δ to rotations. ■

For translational velocity, after tedious computation, the interaction matrix related to Δ can be written as (20), shown at the bottom of the page, and after further computations, it can be shown that

$$\mathbf{L}_{\Delta_v} = \mathbf{L}_{\Delta_{v1}} + \mathbf{L}_{\Delta_{v2}} \tag{21}$$

where

$$\begin{cases} \mathbf{L}_{\Delta_{v_1}} = \dfrac{\mathcal{X}_{31}^\top [\mathcal{X}_{21}]_\times^\top \left[\boldsymbol{\alpha}^\top \mathcal{X}_{21} [\mathcal{X}_{31}]_\times + \boldsymbol{\alpha}^\top \mathcal{X}_{31} [\mathcal{X}_{12}]_\times\right]}{4\Delta} \\ \mathbf{L}_{\Delta_{v2}} = \mathcal{X}_{31}^\top [\mathcal{X}_{21}]_\times^\top \left[\boldsymbol{\alpha}^\top \mathcal{X}_{s_1} [\mathcal{X}_{32}]_\times \mathcal{X}_{s_1} \mathcal{X}_{s_1}^\top \right. \\ \qquad\qquad + \boldsymbol{\alpha}^\top \mathcal{X}_{s_2} [\mathcal{X}_{13}]_\times \mathcal{X}_{s_2} \mathcal{X}_{s_2}^\top \\ \qquad\qquad \left. + \boldsymbol{\alpha}^\top \mathcal{X}_{s_3} [\mathcal{X}_{21}]_\times \mathcal{X}_{s_3} \mathcal{X}_{s_3}^\top \right] / (4\Delta). \end{cases}$$

In practice, $\mathbf{L}_{\Delta_v}$ depends strongly on $\mathbf{L}_{\Delta_{v_1}}$ because the numerator of $\mathbf{L}_{\Delta_{v_2}}$ is a polynomial of point projections with a higher order than the numerator of $\mathbf{L}_{\Delta_{v_1}}$.

III. Feature Choice

In this section, we detail our choice of image features. First, we will explain how to obtain features to control the translational DOFs with interaction matrices that are almost constant with respect to variation in object depth. Then, a vector of features to control all six DOFs will be proposed.

A. Variation of the Interaction Matrix With Respect to Camera Pose

As mentioned above, one of the goals of this study is to decrease the system nonlinearity and coupling by selecting adequate features. The invariance property, for example, results in some interaction-matrix entries being 0, thus removing coupling between DOF as well as being constant during the servoing task. However, other entries depend on the camera pose, as will be shown next. It will be also shown that the feature choice $I_t = 1/\sqrt{I_1}$, and $s_\Delta = 1/\sqrt{\Delta}$ leads to interaction matrices that are almost constant with respect to the object-depth variation.

$$\mathbf{L}_{\Delta_v} = \frac{\mathcal{X}_{21}^\top [\mathcal{X}_{31}]_\times^\top [\mathcal{X}_{31}]_\times \boldsymbol{\alpha}^\top (-\mathcal{X}_{21}\mathbf{I}_3 + \mathcal{X}_{s_2}\mathcal{X}_{s_2}\mathcal{X}_{s_2}^\top - \mathcal{X}_{s_1}\mathcal{X}_{s_1}\mathcal{X}_{s_1}^\top)}{4\Delta} + \frac{\mathcal{X}_{31}^\top [\mathcal{X}_{21}]_\times^\top [\mathcal{X}_{21}]_\times \boldsymbol{\alpha}^\top (-\mathcal{X}_{31}\mathbf{I}_3 + \mathcal{X}_{s_3}\mathcal{X}_{s_3}\mathcal{X}_{s_3}^\top - \mathcal{X}_{s_1}\mathcal{X}_{s_1}\mathcal{X}_{s_1}^\top)}{4\Delta} \tag{20}$$

1) Variation With Respect to Rotational Motion: Let us consider two frames $\mathcal{F}_1$ and $\mathcal{F}_2$ related to the unit sphere with different orientations ($^1\mathbf{R}_2$ is the rotation matrix between the two frames) but with the same center. In this case, the value of I_t is the same for the two frames, since it is invariant to rotation. Let $\mathcal{X}_s$ and $\mathcal{X}'_s = {}^2\mathbf{R}_1\mathcal{X}_s$ be the coordinates in the frame $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, of a projected point. Let us consider a function, which is invariant to rotations $f(\mathcal{X}_1, \ldots, \mathcal{X}_N)$ that can be computed from the coordinates of N points onto the unit sphere. The invariance condition between the frames $\mathcal{F}_1$ and $\mathcal{F}_2$ can be written as

$$f(\mathcal{X}'_1, \ldots, \mathcal{X}'_N) = f({}^2\mathbf{R}_1\mathcal{X}_1, \ldots, {}^2\mathbf{R}_1\mathcal{X}_N) = f(\mathcal{X}_1, \ldots, \mathcal{X}_N). \tag{22}$$

The interaction matrix that links the variation of the function f with respect to translational velocities can be obtained as

$$\mathbf{L}_{\mathbf{f}_\mathrm{v}} = \frac{\partial f(\mathcal{X}_1 + \mathbf{T}, \ldots, \mathcal{X}_N + \mathbf{T})}{\partial \mathbf{T}} \tag{23}$$

where $\mathbf{T}$ is a small translational motion vector. Let us now apply this formula for the camera pose defined by the frame $\mathcal{F}_2$

$$\begin{aligned}\mathbf{L}'_{\mathbf{f}_\mathrm{v}} &= \frac{\partial f(\mathcal{X}'_1 + \mathbf{T}, \ldots, \mathcal{X}'_N + \mathbf{T})}{\partial \mathbf{T}} \\ &= \frac{\partial f({}^2\mathbf{R}_1\mathcal{X}_1 + \mathbf{T}, \ldots, {}^2\mathbf{R}_1\mathcal{X}_N + \mathbf{T}))}{\partial \mathbf{T}}\end{aligned} \tag{24}$$

from which we obtain

$$\mathbf{L}'_{\mathbf{f}_\mathrm{v}} = \frac{\partial f\left({}^2\mathbf{R}_1(\mathcal{X}_1 + {}^1\mathbf{R}_2\mathbf{T}), \ldots, {}^2\mathbf{R}_1(\mathcal{X}_N + {}^1\mathbf{R}_2\mathbf{T})\right)}{\partial \mathbf{T}}. \tag{25}$$

Combining with the rotational invariance condition (22), we obtain

$$\mathbf{L}'_{\mathbf{f}_\mathrm{v}} = \frac{\partial f(\mathcal{X}_1 + {}^1\mathbf{R}_2\mathbf{T}, \ldots, \mathcal{X}_N + {}^1\mathbf{R}_2\mathbf{T})}{\partial \mathbf{T}} \tag{26}$$

which leads to

$$\mathbf{L}'_{\mathbf{f}_\mathrm{v}} = \frac{\partial f(\mathcal{X}_1 + \mathbf{T}', \ldots, \mathcal{X}_N + \mathbf{T}')}{\partial \mathbf{T}'} \frac{\partial \mathbf{T}'}{\partial \mathbf{T}} \tag{27}$$

where $\mathbf{T}' = {}^1\mathbf{R}_2\mathbf{T}$. Finally, combining with (23) yields

$$\mathbf{L}'_{\mathbf{f}_\mathrm{v}} = \mathbf{L}_{\mathbf{f}_\mathrm{v}}\ {}^1\mathbf{R}_2. \tag{28}$$

This result was expected since applying a translational velocity $\mathbf{v}_1$ to the frame $\mathcal{F}_1$ is equivalent to applying a translational velocity to the frame $\mathcal{F}_2$ but taking into account the change of frame ($\mathbf{v_2} = {}^\mathbf{2}\mathbf{R_1}\mathbf{v_1}$). This variation is thus natural—the translational velocity applied to the camera frame depends on its orientation. Finally, this result shows that rotational motions do not change the rank of the interaction matrix of the features used to control the translational DOFs. In other words, the rotational motions do not introduce singularities to the interaction matrix and any rank change of the latter depends only on the translational motion.

2) Variation of the Interaction Matrix With Respect to Depth: Constant interaction-matrix entries are a desirable property and mean that the corresponding features depend linearly of the corresponding DOF. It was shown in [17] and [26] that for good z-axis closed-loop behavior in IBVS, one should choose image features that scale as $s \sim Z$ (where Z is the object depth) so that the variation of their corresponding interaction matrices with respect to depth is zero. In the case where the object is defined by an image region, the following feature has been proposed to control the motion along and around the optical axis [6], [26]:

$$s_r = \frac{1}{\sqrt{m_{00}}}$$

where m_{00} is the moment of order 0 (i.e., object surface in the image) using the conventional perspective projection model. In the case where the object is defined by a set of discrete points, the selected optimal feature for rotation was

$$s_d = \frac{1}{\sqrt{(\mu_{20} + \mu_{02})}} \tag{29}$$

where μ_{ij} are the central moments computed from a set of discrete points; for more details, see [26]. Unfortunately, s_r and s_d only provide invariance to rotations around the optical axis and not to all 3-D rotations.

For this reason, $I_t = 1/\sqrt{I_1}$, and $s_\Delta = 1/\sqrt{\Delta}$ will be used instead of s_d and s_r, respectively. To explain this choice, let us first determine how the polynomial invariant I_1 behaves for increasing Z by considering each of its terms. Let us consider the definition of the projection onto the unit sphere

$$\begin{cases} x_s = \dfrac{X}{\sqrt{X^2 + Y^2 + Z^2}} \\ y_s = \dfrac{Y}{\sqrt{X^2 + Y^2 + Z^2}} \\ z_s = \dfrac{Z}{\sqrt{X^2 + Y^2 + Z^2}}. \end{cases} \tag{30}$$

From (30), we can see that if the depth Z increases (assuming $X \ll Z$ and $Y \ll Z$), the point-projection coordinates have the following behaviors with respect to depth: $x_s \sim 1/Z$, $y_s \sim 1/Z$, and $z_s \sim 1$. It follows that $m_{200} = \sum_{h=1}^N x_{s_h}^2 \sim 1/Z^2$, $m_{020} = \sum_{h=1}^N y_{s_h}^2 \sim 1/Z^2$, $m_{110} = \sum_{h=1}^N x_{s_h} y_{s_h} \sim 1/Z^2$, $m_{101} = \sum_{h=1}^N x_{s_h} z_{s_h} \sim 1/Z$, $m_{011} = \sum_{h=1}^N y_{s_h} z_{s_h} \sim 1/Z$, and $m_{002} = \sum_{h=1}^N z_{s_h}^2 \sim N$. Neglecting the terms that depend on $1/Z^4$, the polynomial can be approximated as

$$I_1 \approx N(m_{200} + m_{020}) - m_{100}^2 - m_{010}^2. \tag{31}$$

Now, we can see that $I_1 \sim 1/Z^2$, and $I_t = 1/\sqrt{I_1} \sim Z$. Note that if the set of points is centered with respect to the optical axis (i.e., $m_{100} = m_{010} = 0$), we have

$$I_1 \approx N(m_{200} + m_{020}). \tag{32}$$

In this case, note the similarity between $I_t = 1/\sqrt{I_1}$ and the features given by (29). In geometric terms, if the set of points is centered with respect to the optical axis, the perspective projections onto the unit sphere and onto the image plane behave in the same way as depth increases. Similarly, it is possible to show that $\Delta \sim 1/z^2$ and $1/\sqrt{\Delta} \sim z$. Examples of interaction-matrix variation with respect to object depth are given in Section IV-A1. Note finally that the interaction matrix related to s_Δ can be

obtained from $\mathbf{L}_\Delta$ as follows:

$$\mathbf{L}_{s_\Delta} = -\frac{1}{2\Delta^{3/2}}\mathbf{L}_\Delta \tag{33}$$

since $\partial s_\Delta/\partial t = -(1/(2\Delta^{3/2}))(\partial\Delta/\partial t)$. The same applies for I_t. From (33), it is clear that the invariance to rotation shown for Δ and I_1 is still valid for s_Δ and I_t.

B. Feature Selection

To control the rotational DOFs, consider the center of gravity of the object's projection onto the unit sphere

$$\mathbf{x}_{\mathbf{s_g}} = (\, x_{s_g}, \quad y_{s_g}, \quad z_{s_g}\,) = \left(\frac{m_{100}}{m_{000}}, \quad \frac{m_{010}}{m_{000}}, \quad \frac{m_{001}}{m_{000}}\right).$$

As in [27], only two coordinates of $\mathbf{x}_{\mathbf{s_g}}$ are useful for control since the point projections belong to the unit sphere making one-coordinate dependent. Recall that the interaction related to $\mathbf{x}_{\mathbf{s_g}}$ is obtained using (10). In order to control rotation around the optical axis, the mean orientation of all line segments in the image is used as a feature. Each segment is built using two different points in an image obtained by reprojection to a conventional perspective plane. More precisely, the segment orientation is defined by

$$\theta = \arctan2\left(\frac{y_2 - y_1}{d}, \frac{x_2 - x_1}{d}\right) \tag{34}$$

where x_i and y_i are the coordinates of the points forming the segment in a classical perspective plane, and $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ is the distance separating them. By deriving (34) and combining the result with the interaction matrix related to the Cartesian-point coordinates, the interaction matrix related to θ is given by [7]

$$L_\theta = [\,\theta_{\mathbf{vx}} \quad \theta_{\mathbf{vy}} \quad \theta_{\mathbf{vz}} \quad \theta_{\boldsymbol{\omega x}} \quad \theta_{\boldsymbol{\omega y}} \quad -1\,] \tag{35}$$

where

$$\begin{cases} \theta_{\mathbf{vx}} = \dfrac{\sin\theta}{d}\left(\dfrac{1}{Z_2} - \dfrac{1}{Z_1}\right) \\ \theta_{\mathbf{vy}} = \dfrac{-\cos\theta}{d}\left(\dfrac{1}{Z_2} - \dfrac{1}{Z_1}\right) \\ \theta_{\mathbf{vz}} = \dfrac{(x_2y_1 - x_1y_2)}{d^2}\left(\dfrac{1}{Z_2} - \dfrac{1}{Z_1}\right) \\ \theta_{\boldsymbol{\omega x}} = \dfrac{x_1y_1y_2 + x_2y_1y_2 - x_2y_1^2 - x_1y_2^2}{d^2} \\ \theta_{\boldsymbol{\omega y}} = \dfrac{x_1y_1x_2 + x_1x_2y_2 - y_2x_1^2 - y_1x_2^2}{d^2} \end{cases}$$

and Z_i is the depth of the 3-D points.

Finally, the invariants to 3-D rotation $I_t = 1/\sqrt{I_1}$, or $s_\Delta = 1/\sqrt{\Delta}$ are considered as features to control the translational DOF. The choice of $I_t = 1/\sqrt{I_1}$ or $s_\Delta = 1/\sqrt{\Delta}$ is discussed in Section IV-E. In practice, when I_t is used, three separate targets such that their centers are noncollinear can be sufficient to control the translational DOFs. In order to ensure the nonsingularity of the interaction matrix, the set of points is divided into four subsets (each subset must contain at least three points). This allows us to obtain four different features to control the three translational DOFs. Recall that if the set is composed of coplanar points, the simple form (11) of the interaction matrix can be used. If the points are noncoplanar, form (10) has to be used. Similarly, when s_Δ is used, four different triangles can be obtained by combining three noncollinear points among a set of at least four points.

IV. Results

In this section, the theoretical results presented above are validated. We will first see how the interaction matrix varies with the depth through two examples. Then, a statistical study of convergence rate using several features is given. Finally, we present the results of a series of experiments with planar and nonplanar targets and two kinds of camera (i.e., conventional and fisheye).

A. Variation of the Interaction Matrix With Camera Pose

1) Variation With Respect to Camera Translation: Fig. 3 shows the variation of the interaction-matrix entries related to Δ and $1/\sqrt{\Delta}$ with respect to translational motion applied to the triangle

$$\mathcal{X} = \begin{bmatrix} -0.15 & -0.15 & 0.3 \\ 0.2598 & -0.2598 & 0 \\ 0.5 & 0.5 & 0.5 \end{bmatrix} \tag{36}$$

defined in the camera frame. From Fig. 3(a) and (d), it can be seen that $L_x = L_{x_1} = L_y = L_{y_1} = 0$, irrespective of object depth (where $\mathbf{L}_\Delta = [L_x,\ L_y,\ L_z,\ 0,\ 0,\ 0]$, and $\mathbf{L}_{s_\Delta} = [L_{x_1},\ L_{y_1},\ L_{z_1},\ 0,\ 0,\ 0]$). In practice, the features Δ and s_Δ depend mainly on the translational motion with respect to the object axis of view (in this example, the axis of view is the camera optical axis). From Fig. 3(d), it can also be seen that L_{z_1} is almost constant and invariant to the object depth. On the other hand, L_z decreases to 0 when the object depth increases [see Fig. 3(a)].

The variation of interaction-matrix entries for translational motion with respect to *x*-axis and *y*-axis motion are shown in Fig. 3(b), (c), (e), and (f). First, it can be seen that *x*-axis translational motion influences mainly the entries corresponding to the *x*-axis and *z*-axis. Similarly, the *y*-axis translational motion influences mainly the entries corresponding to the *y*-axis and *z*-axis. Furthermore, variation of the interaction-matrix entries for the *x*-axis and *y*-axis translational motion are more uniform for s_Δ than for Δ.

As a second example, Fig. 4 shows the variations of the interaction-matrix entries of I_1 and $1/\sqrt{I_1}$ with respect to translational motion along the *z*-axis applied to the four coplanar points defined in the camera frame

$$\mathbf{X_o} = \begin{pmatrix} -0.3258 & -0.0811 & 0.1487 & 0.2583 \\ -0.0458 & 0.1470 & -0.1052 & 0.0039 \\ 1 & 1 & 1 & 1 \end{pmatrix}. \tag{37}$$

The set of points has been chosen to be approximatively centered with respect to the *z*-axis ($m_{100} \approx 0$, and $m_{010} \approx 0$). For this

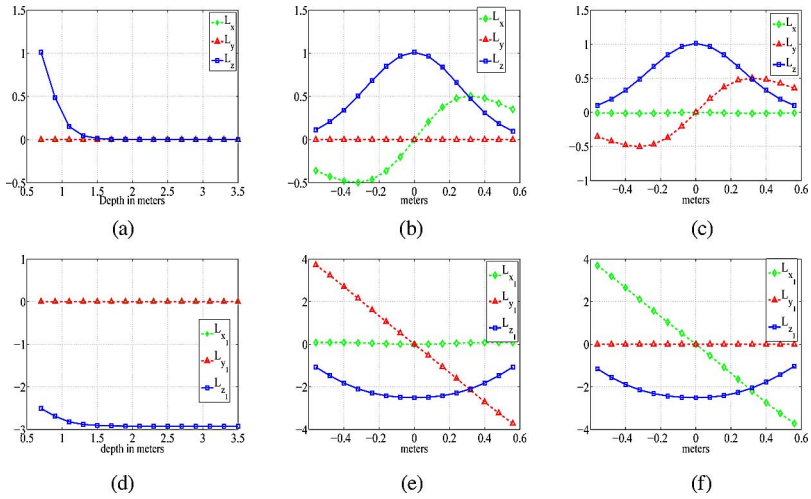


Fig. 3. Results obtained for $s = \Delta$. (a) Variation with respect to depth, (b) variation with respect to x-axis translation, and (c) variation with respect to y-axis translation. Results obtained for $s_\Delta = 1/\sqrt{\Delta}$. (d) Variation with respect to depth, (e) variation with respect to x-axis translation, and (f) variation with respect to y-axis translation.

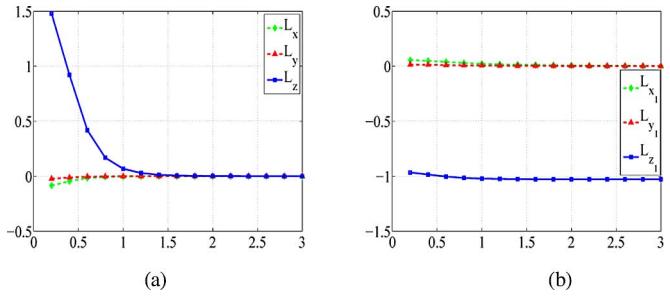


Fig. 4. Variation of the interaction matrix with object depth (in meters). (a) Results for I_1 and (b) results for I_t.

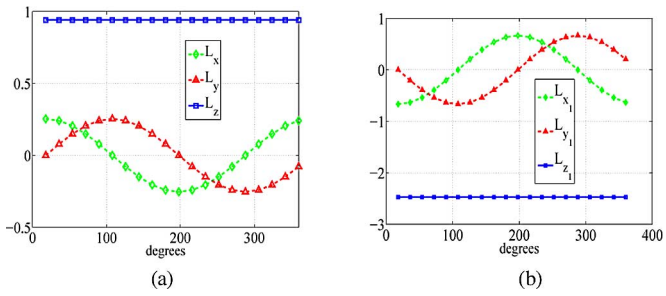


Fig. 5. Variation of the interaction-matrix entries with respect to a rotation around the optical axis. (a) Results for $s = \Delta$ and (b) results for s_Δ.

reason, it can be seen that $L_x \approx L_{x_1} \approx L_y \approx L_{y_1} \approx 0$ ($\mathbf{L}_{I_1} = [L_x,\ L_y,\ L_z,\ 0,\ 0,\ 0]$, and $\mathbf{L}_{I_t} = [L_{x_1},\ L_{y_1},\ L_{z_1},\ 0,\ 0,\ 0]$). In practice, the features I_1 and $1/\sqrt{I_1}$ also depend mainly on the translational motion with respect to the object axis of view. From Fig. 4(a) and (b), it can also be seen that L_{z_1} is almost constant and largely invariant to the object depth, while L_z decreases to 0 when the object depth increases.

2) Variation With Respect to the Camera Orientation: To illustrate the results presented in Section III-A1, we have computed the variations of the interaction-matrix entries related to Δ and $1/\sqrt{\Delta}$ with respect to rotation around the optical axis applied to the triangle defined by the following 3-D coordinates in the camera frame:

$$\mathcal{X} = \begin{bmatrix} -0.05 & -0.05 & 0.4 \\ 0.2598 & -0.2598 & 0 \\ 0.5 & 0.5 & 0.5 \end{bmatrix}. \tag{38}$$

These results are shown in Fig. 5. The curves corresponding to the entries L_z and L_{z_1} show that they are constant. This was expected, since a rotational motion around the optical axis does not change the orientation of this axis. Thus, the variation of the selected features with respect to rotational motion around the optical axis remains constant. From the same figure, it can also be seen that the variation of the other entries are sinusoidal functions. In fact, as shown in Section III-A1, the interaction matrix after a rotational motion is the product of this matrix by the rotation matrix.

B. Simulation Results Using a Nonplanar Set of Points

In these simulations, the target comprises four noncoplanar points. The desired position corresponds to the following 3-D point coordinates defined in the camera frame:

$$\mathbf{X_d} = \begin{pmatrix} 0 & -0.2 & 0 & 0.2 \\ 0.2 & 0 & -0.2 & 0 \\ 0.9 & 1. & 1 & 1.2 \end{pmatrix}. \tag{39}$$

Two different features set are tested to control the translational motion: $s_\Delta = 1/\sqrt{\Delta}$, and $I_t = 1/\sqrt{I_1}$. From the four points, four different triangles can be obtained. For each triangle, the invariants s_Δ and I_t are used.

In this simulation, we show the advantage of using s_Δ or I_t instead of directly using Δ and I_1. For this purpose, the translational motion $\mathbf{t_0} = (0.2\ \mathrm{m},\ 0.3\ \mathrm{m},\ 0.6\ \mathrm{m})$ has been considered

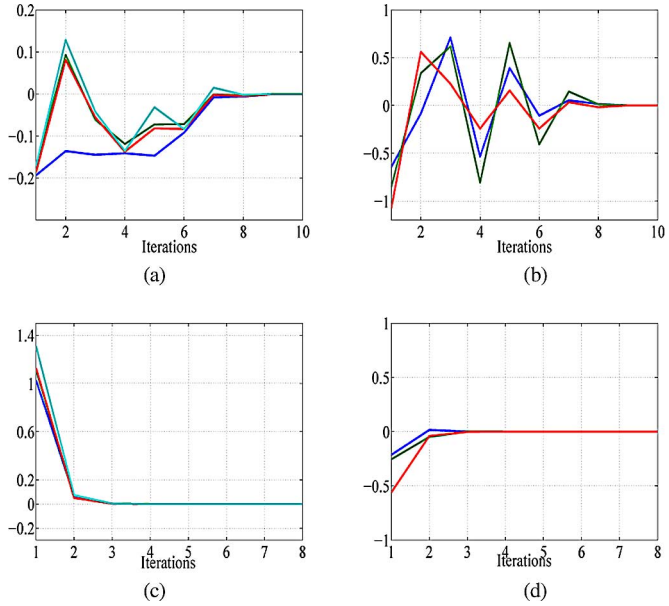


Fig. 6. Results obtained using I_1. (a) Feature errors and (b) velocities (in meters per second); results obtained using I_t. (c) Feature errors and (d) velocities (in meters per second).

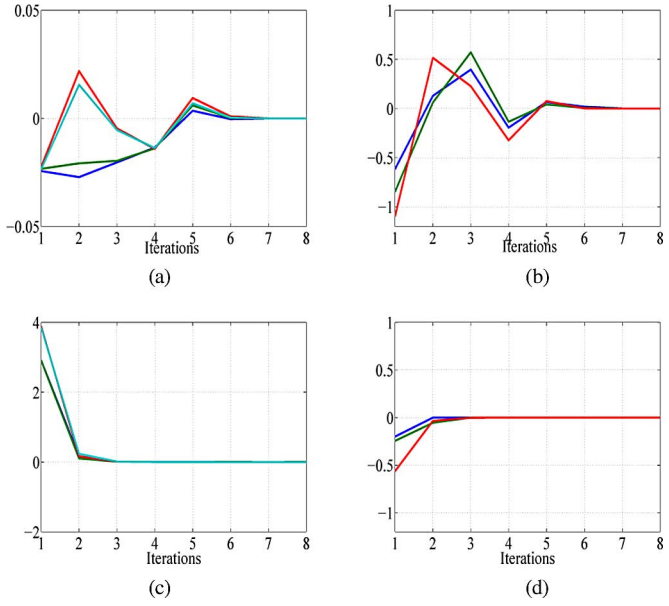


Fig. 7. Results obtained using Δ. (a) Feature errors and (b) velocities (in meters per second); results obtained using s_Δ. (c) Feature errors and (d) velocities (in meters per second).

between the desired and the initial camera poses. The scalar-velocity gain in the control law has been set to $\lambda = 1$. If the system were completely linear, convergence would be obtained in only one iteration. The nonlinearity of the system has the effect of damping or magnifying the camera velocities. The results obtained using I_t and I_1 are shown in Fig. 6. From Fig. 6(a) and (b), we observe oscillations for both the feature error and the velocities obtained using I_1 before converging (after nine iterations). On the other hand, a fast convergence is obtained using I_t without oscillation (after only two iterations, the system has almost converged). This shows that using I_t, the system has behaved as an almost-linear system. The results obtained using s_Δ and Δ are shown in Fig. 7. They confirm those obtained using I_t and I_1—the system converges faster using s_Δ rather than Δ.

In the second simulation, a generic motion combining the rotational motion $\boldsymbol{\theta}\mathbf{u} = (-7.90^\circ,\ 23.70^\circ,\ 158.0^\circ)$ and the translational motion $\mathbf{t}_1 = (0\ \text{m},\ 0.3\ \text{m},\ 1\ \text{m})$ is considered.

The results obtained using s_Δ to control translation are shown in Fig. 8(a)–(c). Despite the large motion, it can be seen that a satisfactory behavior is obtained for the feature errors [see Fig 8(a)]. Similar satisfactory behaviors are simultaneously obtained for the velocities [see Fig. 8(b) and (c)]. Furthermore, from the obtained results (see the plot corresponding to the translational velocities), it can be seen that one of the velocities (i.e., translation with respect to x-axis) is null, which means that the considered displacements with respect to the y-axis and the z-axis produce no velocity with respect to the x-axis. This confirms the decoupling properties of our control. Finally, the results obtained using I_t to control the translational motions are shown in Fig. 8(d)–(f). From these figures, it can be seen that the behavior is satisfactory and almost identical to the one obtained using s_Δ.

C. Convergence Rate and Robustness

In this part, comparison of the visual servoing using the features I_t, s_Δ, and the Cartesian image-point coordinates is made. More precisely, the convergence rate for random initial and desired poses are computed using the different features in the case of perfect data as well as in the case when errors on camera calibration and noisy image points occur. The following setup has been used.

1) Two objects composed, respectively, of four points forming a square [see (40)] and six points generated randomly [see (41)] have been considered (see, respectively, Fig. 9(a) and (b))

$$X_1 = \begin{bmatrix} -0.4 & 0.4 & -0.4 & 0.4 \\ -0.4 & -0.4 & 0.4 & 0.4 \\ 1. & 1. & 1. & 1 \end{bmatrix} \tag{40}$$

$$X_2 = \begin{bmatrix} -0.31 & 0.68 & -0.00 & -0.69 & 0.42 & -0.09 \\ 0.07 & -0.09 & -0.41 & 0.40 & -0.28 & 0.32 \\ 0.96 & 1.04 & 0.99 & 1.01 & 0.99 & 0.95 \end{bmatrix}. \tag{41}$$

2) A conventional camera model with focal $F = 800$ and principal-point coordinates $u = v = 400$ pixels has been used to compute the image-point coordinates.
3) The interaction matrix corresponding to the current position is used in the control law (4) to compute the camera displacement (i.e., $\widehat{\mathbf{L}}_\mathbf{s} = \mathbf{L}_\mathbf{s}$) and the scalar λ has been set to 0.1.
4) The initial and the desired camera poses have been generated randomly as follows.
   a) A total of 600 random translational motions $\mathbf{t} = (1.5\,\sigma_1,\ 1.5\,\sigma_2,\ 1.7\sigma_3)$ are first applied to the

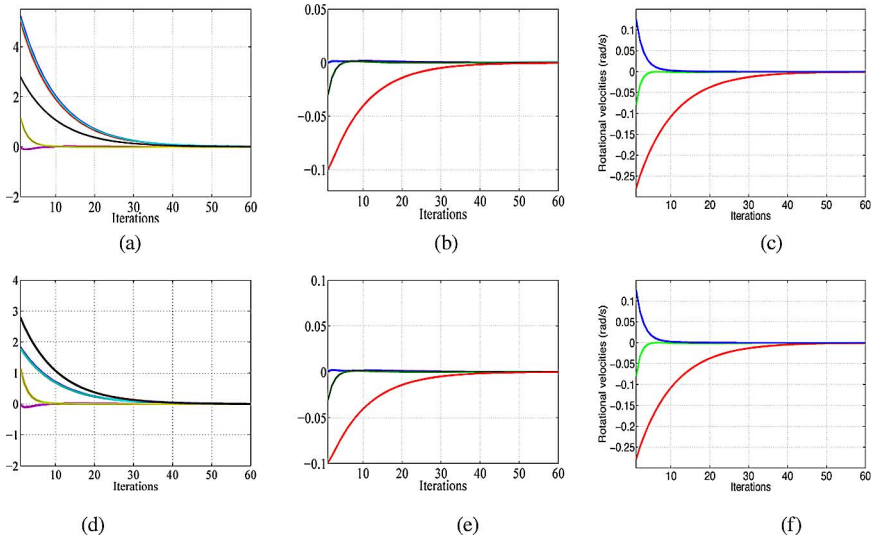


Fig. 8. Results for large general motion using s_Δ. (a) Feature errors, (b) translational velocities (in meters per second), and (c) rotational velocities (in radians per second). Results for large general motion using I_t. (d) Feature errors, (e) translational velocities (in meters per second), and (f) rotational velocities (in radians per second)).

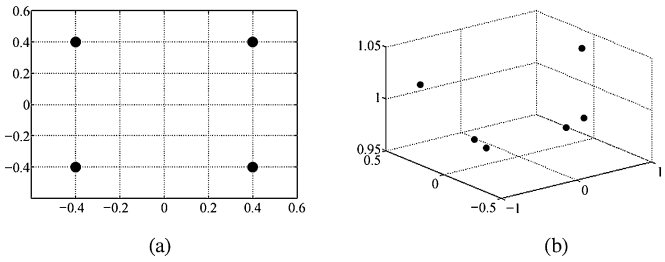


Fig. 9. (a) Square-point coordinates and (b) object composed of six points generated randomly.

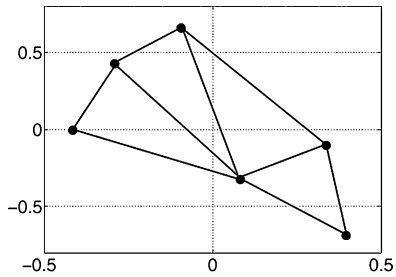


Fig. 10. Selected four triangles to compute the invariant s_Δ.

TABLE I
CONVERGENCE RATE USING THE SQUARE

Features	s_Δ	I_t	Point coordinates
Perfect data	100%	74.4%	62.1%
Noisy	99.83%	73.83%	62.33%

TABLE II
CONVERGENCE RATE USING THE OBJECT COMPOSED OF SIX NONCOPLANAR POINTS

s_Δ (4 triangles)	s_Δ (all triangles)	I_t (all triangles)	Point coordinates
99.8%	85.6%	86%	78.8%

point coordinates defined in the object frame, where σ_1 and σ_2 are random numbers chosen from a normal distribution with mean 0 and standard deviation 1, and σ_3 is a random number chosen from a uniform distribution on the interval $[0.0\ 1.0]$.

b) The rotational motion is chosen such that the points coordinates belongs to the image limits $[1\ 800; 1\ 800]$. Further, the rotational motion with respect to the optical axis can range randomly between $[0\ \ 2\pi]$.

The results obtained using the square are given in Table I. The first line of Table I gives the percentage of convergence for the case when perfect data are used. It can be noted that for the considered desired and initial poses, s_Δ allows obtaining the highest convergence rate, followed by I_t, and then the Cartesian-point coordinates. The results obtained with errors on camera parameters (10% errors on focal length and 20 pixels error on principal-point coordinates) are similar and show the superiority of the feature s_Δ. Furthermore, the convergence rate does not suffer from the errors on camera parameters and noise.

Let us now discuss the results obtained using the object composed of six noncoplanar points. Two cases have been considered for s_Δ. In the first case, the feature vector is defined by the invariants s_Δ computed for all possible ten triangles obtained by combining three different points, while in the second case, only four selected triangles are used. The four triangles are selected such that their centers of gravity are not collinear and the farthest possible from each other to allow a good conditioning of the interaction matrix (see Fig. 10). Table II gives the convergence rate using, respectively, the features s_Δ (using four triangles), s_Δ (using all triangles), I_t (using all triangles) and, finally, the Cartesian-point coordinates. The obtained results confirm those obtained using the square—the use of s_Δ allows obtaining a better convergence rate that the other features. Furthermore, the

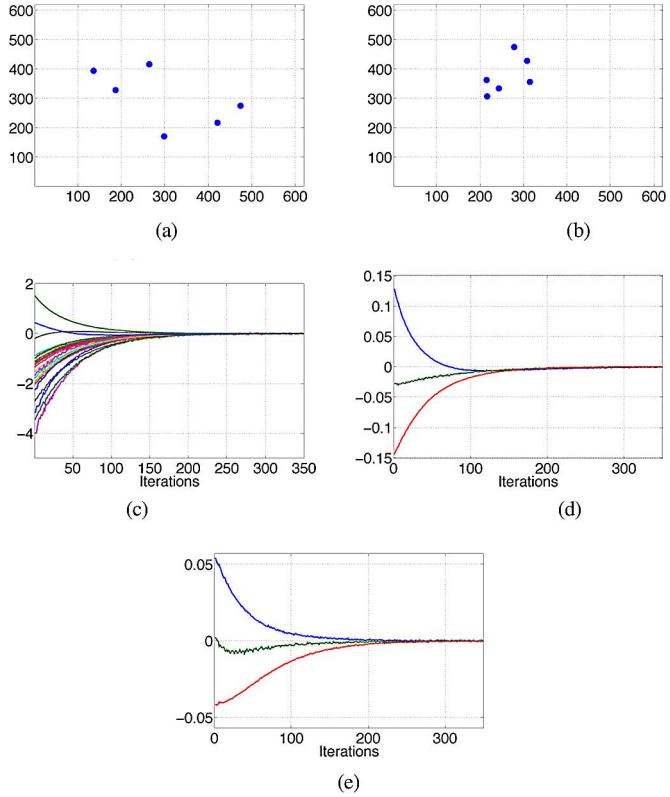


Fig. 11. Effect of noisy data on the computed velocities. (a) Desired image points, (b) initial image points, (c) feature errors, (d) rotational velocities (in radians per second), and (e) translational velocities (in meters per second).

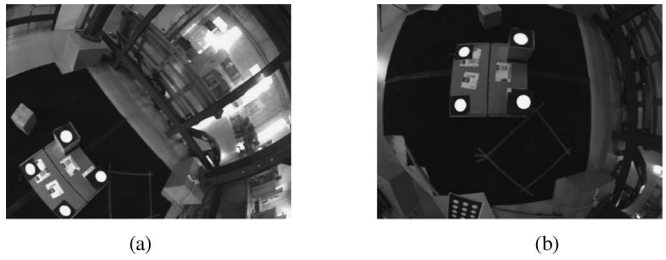


Fig. 12. Results using fisheye camera and a noncoplanar set of four points. (a) Initial image and (b) desired image.

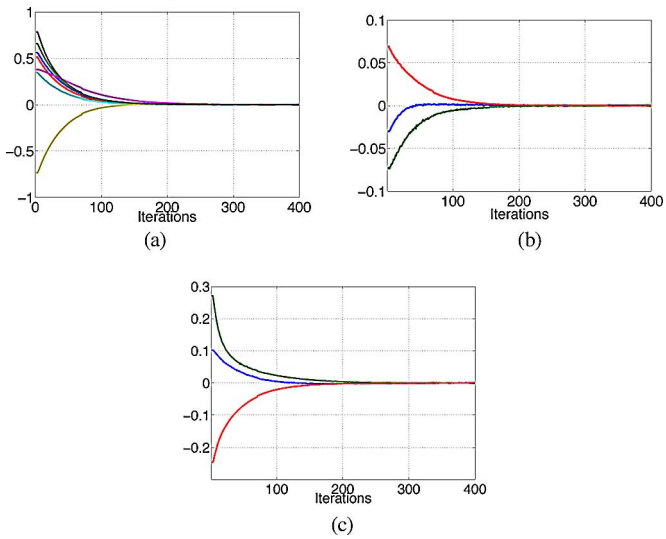


Fig. 13. Results using fisheye camera and a noncoplanar set of four points (generic motion using s_Δ). (a) Feature errors, (b) translational velocities (in meters per second), and (c) rotational velocities (in radians per second).

convergence rate when only four selected triangles are used to compute s_Δ increased significantly; using only four features to control the translational motions allowed avoiding local minima.

Finally, the statistical studies of convergence rate show that the majority of the cases when the camera does not converge to its desired position are mainly due to local minima. The cases of complete divergence usually happen because of errors on the used interaction-matrix values.

In these last simulation results, we test the effect of the noise on image-point coordinates on the computed velocities using s_Δ to control the translations. For this purpose, the object defined by (41) and an omnidirectional camera with focal $F = 500$, principal-point coordinates $u = v = 300$ pixels, and camera parameter $\xi = 2$ have been considered. The desired and the initial image points are shown, respectively, in Fig. 11(a) and (b). A generic motion that combines the rotational motion $\boldsymbol{\theta}\mathbf{u} = (-27.88^\circ,\ 41.82^\circ,\ 69.70^\circ)$ and the translational motion $\mathbf{t} = (-0.4\ \text{m},\ 0.2\ \text{m},\ 0.7\ \text{m})$ is considered between the initial and the desired camera positions. Further, white noise with standard deviation equal to 0.5 pixels has been added to the image-point coordinates. The obtained behaviors for the feature errors, the rotation velocities, and the translational ones are shown, respectively, in Fig. 11(c)–(e). From these figures, it can be noted that the computed velocities are once again satisfactory and robust to the considered noise on image points.

D. Experimental Results Using a Conventional and a Fisheye Cameras

In the following, a series of experiments using conventional and fisheye cameras are presented. The parameters of the classical perspective camera used in these experiments are ($\xi = 0$, principal-point coordinates $[u_x = 534.8$ pixels, $u_u = 353.4$ pixels$]$, the focal $[F_x = 960$ pixels, $F_y = 959$ pixels$]$). Those of the fisheye camera used in these experiments are ($\xi = 1.71$, principal-point coordinates $[u_x = 315.61$ pixels, $u_u = 243.05$ pixels$]$, the focal $[F_x = 722.91$ pixels, $F_y = 721.65$ pixels$]$). Two different objects that are composed, respectively, of four coplanar points and four noncoplanar points will be considered. The calibration of the two cameras was performed using the toolbox provided by Mei and Rives [20].

1) Results Using a Set of Four Noncoplanar Points: In this part, an experimental result using fisheye camera and generic motion involving large rotational and translational motions is presented. Both the behaviors of the control law using s_Δ and I_t are tested. The interaction matrix computed for the current camera position is used in the control law. The images corresponding to the camera initial and desired poses are shown in Fig 12. The results obtained using s_Δ are shown in Fig. 13, and those obtained using I_t are shown in Fig. 14. The plots show

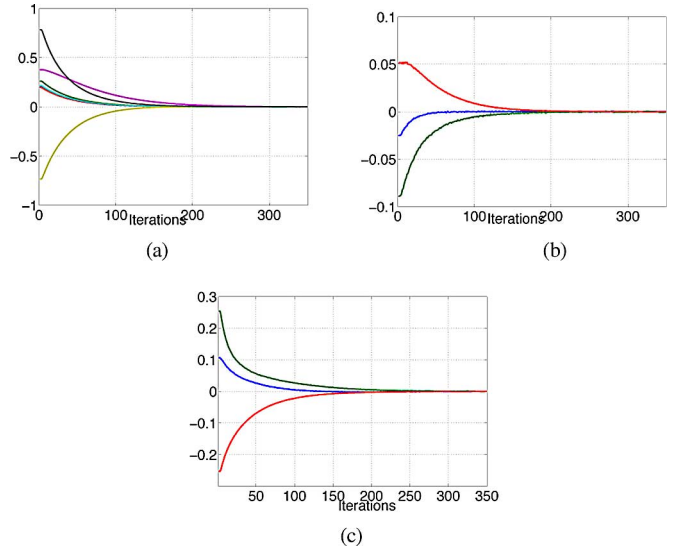


Fig. 14. Results using fisheye camera and a noncoplanar set of four points (generic motion using I_t). (a) Feature errors, (b) translational velocities (in meters per second), and (c) rotational velocities (in degrees per second).

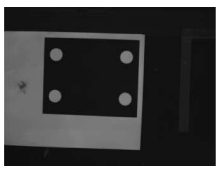


Fig. 15. Desired image using a conventional camera.

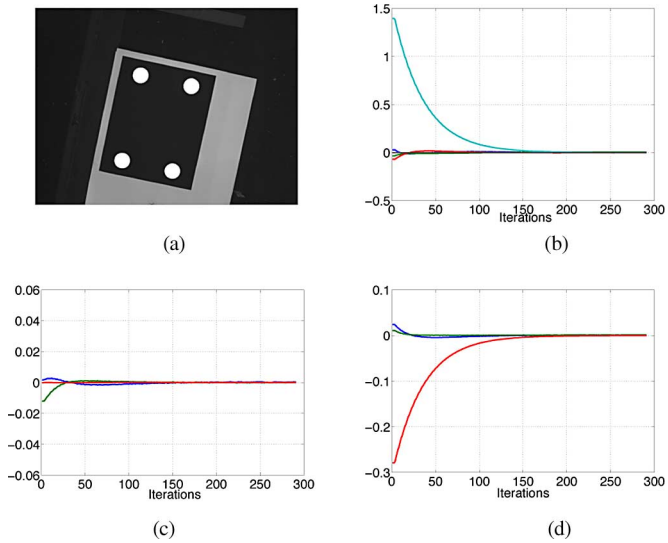


Fig. 16. Results for pure-rotation motion (80°) using conventional camera. (a) Initial image, (b) feature errors, (c) translational velocities (in meters per second), and (d) rotational velocities (in radians per second).

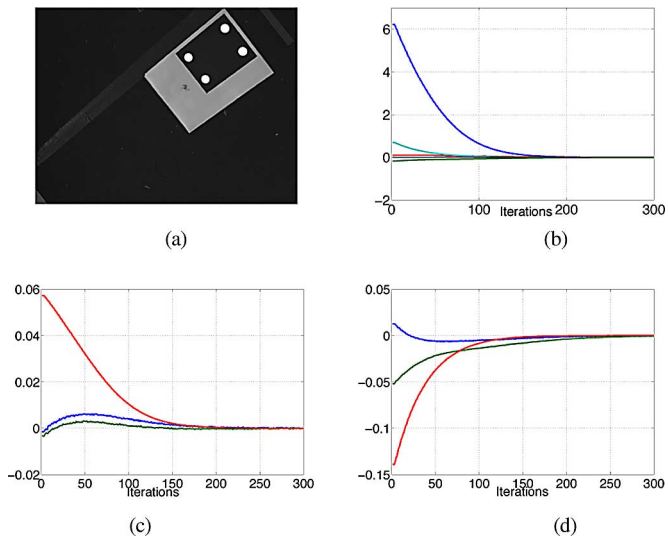


Fig. 17. Results for complex motion using conventional camera. (a) Initial image, (b) feature errors, (c) translational velocities (in meters per second), and (d) rotational velocities (in radians per second).

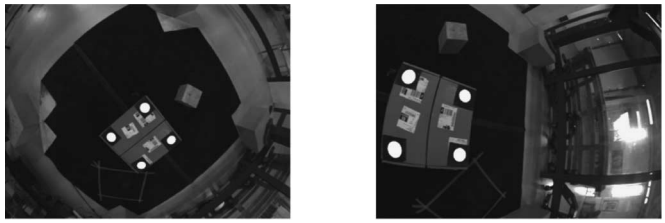
that the error in feature value as well as the velocities converge to 0 in very satisfactory ways using s_Δ and I_t.

2) Results Using a Set of Four Coplanar Points: In this part, experiments using the feature s_Δ to control the translational motions are presented.

a) Results using a perspective camera: In the following experiment, the interaction matrices are computed using the current values of the points in the image and constant approximated desired point depths. The first experiment involves pure rotational motion around the camera optical axis (i.e., 80°). The desired and the initial images are shown, respectively, in Figs. 15 and 16(a). Four combinations of triangles obtained from the points defining the target are used to control translational motion. The results are shown in Fig. 16(b)–(d). Fig. 16(b) shows that a nice decrease of feature error is obtained. Furthermore, since the considered translational motion is null, the translational velocity computed using the invariants to rotations are almost null [see Fig. 16(b)]. The small translational velocities are due to the weak calibration of the camera. Fig.16(d) shows good behavior for rotational motions as well.

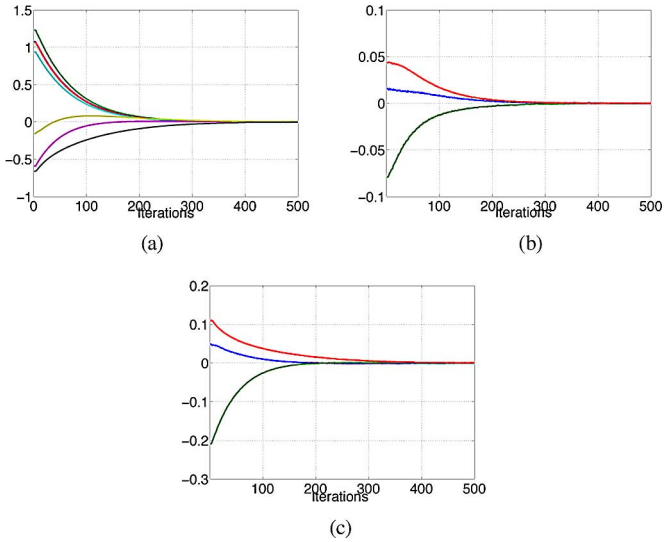
The second experiment using a perspective camera involves a complex motion between the desired and the initial camera positions, as shown in Figs. 15 and 17(a), respectively. From Fig. 17(b), we note that the feature error behaves in a very satisfactory way, despite the errors in camera calibration and points depth (the point depths are not computed at each iteration). Satisfactory behaviors are also obtained for translational and rotational velocities [see Fig. 17(c) and (d)]. Indeed, nice decreases of the feature errors as well as of the velocities are obtained.

b) Results using a fisheye camera: In this experiment, a generic displacement involving translational and rotational motion has been considered. The images corresponding to the initial and the desired camera positions are shown, respectively, in Fig. 18(a) and (b). The displacement to perform is very large, and the camera-desired orientation with respect to the object plane is approximately equal to 60°. In the control law, the interaction matrix computed for the current camera position is



(a) (b)

Fig. 18. Results for four coplanar points (i.e., generic motion using s_Δ). (a) Initial image and (b) desired image.



(a) (b) (c)

Fig. 19. Results for four coplanar points (i.e., generic motion using s_Δ). (a) Feature errors, (b) translational velocities (in meters per second), and (c) rotational velocities (in radians per second).

used. The obtained results are shown in Fig. 19. From the plots, it can be noted that very satisfactory decrease of the error on features as well as for the velocities is obtained. This shows that the orientation of the camera in its desired position does not impact the behavior of the proposed control scheme. This result was expected since a rotation motion of the camera only introduces a rotation of the interaction matrix related to the invariance of the rotations.

E. Discussion

We have shown that two different features can be obtained from a set of points to control the translational DOFs: the features s_Δ obtained from the surfaces of the projection of triangles onto a unit sphere and the polynomials invariant to rotations I_t. The results that we have obtained show that the two invariants provide good decoupling properties.

This raises the question "*which is the best feature to use?*". The experimental results and, more precisely, the convergence rate using two different objects composed, respectively, of four and six points have shown the superiority of the control using s_Δ. However, in the case when the target comprises many points, the number of possible combinations of three points increases combinatorially. This means that the size of the feature vector increases and reaching local minima becomes possible. In order to avoid these local minima, we can consider using only the four *best* triangles, e.g., those triangles that result in the best conditioning of the interaction matrix for the desired position. However, since not all points would be used for servoing, the robustness to noise might decrease. In order to improve the robustness to noise, all the possible combinations of points can be used near the desired position.

Alternatively, the invariant polynomial could be computed from more than three points. Therefore, if the set of points is divided into only four different subsets, we can obtain an almost-minimal representation (i.e., four features to control the three translational DOFs).

The projection surface also has the advantage of being suitable not only for points but for closed contours as well. The surface of the object's projection onto a sphere is simply the moment of order 0, and it can be computed using the general formula

$$m_{s_{i,j,k}} = \int\int_{\text{region}} x_s^i y_s^j z_s^k \, ds. \tag{42}$$

The surface is a generic descriptor that can be computed from an image region defined by a closed and complex contour or simply by a polygonal curve. As for the surfaces obtained by projection of triangles, it is also possible to show that the feature $1/\sqrt{m_{000}}$ has the same properties as s_Δ and I_t with respect to object-depth variation. The decoupled control proposed for objects defined by a set of points can thus be extended straightforwardly to the case where several matched planar contours are available in the scene. More precisely, at least three planar contours are required to control the three translational DOFs.

In order to control the rotational DOFs, features similar to those proposed for the case of objects defined by a set of point can also be used. Indeed, the center-of-gravity coordinates, which are defined as $\mathbf{x}_{\mathbf{s}_\mathbf{g}} = (m_{100}/m_{000}, m_{010}/m_{000}, m_{001}/m_{000})$, can also be computed using the moment definition (42). Finally, in order to control the rotation around the optical axis, the object orientation in the image ($\theta = (1/2)\arctan(2\mu_{11}/(\mu_{20} - \mu_{02}))$) can be used, as given in [26]. The interaction matrix related to region-based moments computed from the projection onto sphere has already been computed in [25]. The analytical formulas of the features as well as their related interaction matrices that are required for the control are available. The extension of the image-based control, which is obtained for a set of points, to the case of multiple matched planar contours is then possible.

V. Conclusion and Future Work

In this paper, a generic decoupled image-based control using the projection onto the unit sphere has been proposed. Invariants to rotation have been used to control translational motion. The proposed decoupled control is valid for all cameras that obey the unified camera model. Further, it is valid for objects defined by at least three planar closed contours or by a set of at least four points. Importantly, the proposed features result in an interaction

matrix whose elements are only weakly dependent on the depth of object points and camera position. Finally, the controller has been experimentally validated and results presented using two kinds of camera: conventional and fisheye. The results show very satisfactory behavior for both 3-D space and the image. Future work will be devoted to extending these results to the pose-estimation problem.

REFERENCES

[1] S. Baker and S. Nayar, "A theory of catadioptric image formation," *Int. J. Comput. Vis.*, vol. 35, no. 2, pp. 175–196, Nov. 1999.

[2] J. Barreto and H. Araujo, "Issues on the geometry of central catadioptric image formation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2001, pp. II-422–II-427.

[3] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The Confluence of Vision and Control* (LNCS Series, no. 237), D. Kriegman, G. Hager, and A. Morse, Eds. New York: Springer-Verlag, 1998, pp. 66–78.

[4] G. Chesi, "Visual servoing path-planning via homogeneous forms and LMI optimizations," *IEEE Trans. Robot.*, vol. 25, no. 5, pp. 281–291, Apr. 2009.

[5] G. Chesi and A. Vicino, "Visual servoing for large camera displacements," *IEEE Trans. Robot.*, vol. 20, no. 4, pp. 724–735, Aug. 2004.

[6] P. I. Corke and S. A. Hutchinson, "A new partitioned approach to image-based visual servo control," *IEEE Trans. Robot. Autom.*, vol. 17, no. 4, pp. 507–515, Aug. 2001.

[7] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. Robot. Autom.*, vol. 8, no. 3, pp. 313–326, Jun. 1992.

[8] J. Feddema and O. Mitchell, "Vision-guided servoing with feature-based trajectory generation," *IEEE Trans. Robot. Autom.*, vol. 5, no. 5, pp. 691–700, Oct. 1989.

[9] D. Fioravanti, B. Allotta, and A. Rindi, "Image based visual servoing for robot positioning tasks," *Meccanica*, vol. 43, no. 3, pp. 291–305, Jun. 2008.

[10] C. Geyer and K. Daniilidis, "Mirrors in motion: Epipolar geometry and motion estimation," *Int. J. Comput. Vis.*, vol. 45, no. 3, pp. 766–773, 2003.

[11] H. Hadj-Abdelkader, Y. Mezouar, P. Martinet, and F. Chaumette, "Catadioptric visual servoing from 3d straight lines," *IEEE Trans. Robot.*, vol. 24, no. 3, pp. 652–665, Jun. 2008.

[12] E. Hall, *Computer Image Processing and Recognition*. New York: Academic, 1979.

[13] T. Hamel and R. Mahony, "Visual servoing of an under-actuated dynamic rigid body system: An image-based approach," *IEEE Trans. Robot. Autom.*, vol. 18, no. 2, pp. 187–198, Apr. 2002.

[14] M. Iwatsuki and N. Okiyama, "A new formulation of visual servoing based on cylindrical coordinates system with shiftable origin," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Lausanne, Switzerland, Oct. 2002, pp. 354–359.

[15] J. T. Lapreste and Y. Mezouar, "A Hessian approach to visual servoing," in *Proc. Int. Conf. Intell. Robots Syst.*, Sendai, Japan, Sep. 28–Oct. 2, 2004, pp. 998–1003.

[16] J. S. Lee, I. Suh, B. J. You, and S. R. Oh, "A novel visual servoing approach involving disturbance observer," in *Proc. IEEE Int. Conf. Robot. Autom.*, Detroit, MI, May 1999, pp. 269–274.

[17] R. Mahony, P. Corke, and F. Chaumette, "Choice of image features for depth-axis control in image-based visual servo control," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Lausanne, Switzerland, Oct. 2002, vol. 1, pp. 390–395.

[18] E. Malis, "Improving vision-based control using efficient second-order minimization techniques," in *Proc. IEEE Int. Conf. Robot. Autom.*, New Orleans, LA, Apr. 2004, vol. 2, pp. 1843–1848.

[19] A. G. Mamistvalov, "n-dimensional moment invariants and conceptual mathematical theory of recognition n-dimensional solids," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 819–831, Aug. 1998.

[20] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3945–3950.

[21] Y. Mezouar and F. Chaumette, "Path planning for robust image-based control," *IEEE Trans. Robot. Autom.*, vol. 18, no. 4, pp. 534–549, Aug. 2002.

[22] P. Rives and J. Azinheira, "Linear structures following by an airship using vanishing points and horizon line in a visual servoing scheme," in *Proc. IEEE Int. Conf. Robot. Autom.*, New Orleans, LA, Apr. 2004, pp. 255–260.

[23] F. Schramm, F. Geffard, G. Morel, and A. Micaelli, "Calibration free image point path planning simultaneously ensuring visibility and controlling camera path," in *Proc. IEEE Int. Conf. Robot. Autom.*, Roma, Italy, Apr. 2007, pp. 2074–2079.

[24] T. Svoboda and T. Pajdla, "Epipolar geometry for central catadioptric cameras," *Int. J. Comput. Vis.*, vol. 49, no. 1, pp. 23–37, Aug. 2002.

[25] O. Tahri, "Utilisation des moments en asservissement visuel et en calcul de pose," Ph.D. dissertation, Univ. Rennes, Rennes, France, 2004.

[26] O. Tahri and F. Chaumette, "Point-based and region-based image moments for visual servoing of planar objects," *IEEE Trans. Robot.*, vol. 21, no. 6, pp. 1116–1127, Dec. 2005.

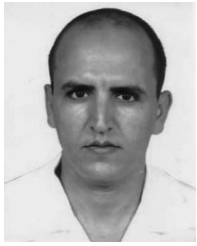
[27] O. Tahri, F. Chaumette, and Y. Mezouar, "New decoupled visual servoing scheme based on invariants from projection onto a sphere," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2008, pp. 3238–3243.

[28] O. Tahri and Y. Mezouar, "On visual servoing based on efficient second order minimization," *Robot. Auton. Syst.*, vol. 58, no. 5, pp. 712–719, May 2010.

[29] O. Tahri, Y. Mezouar, F. Chaumette, and P. Corke, "Generic decoupled image-based visual servoing for cameras obeying the unified projection model," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp. 1116–1121.

[30] R. T. Fomena and F. Chaumette, "Visual servoing from spheres using a spherical projection model," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 2080–2085.

[31] L. Weiss, A. C. Sanderson, and C. P. Neuman, "Dynamic sensor-based control of robots with visual feedback," *IEEE J. Robot. Autom.*, vol. RA-3, no. 5, pp. 404–417, Oct. 1987.



Omar Tahri was born in Fez, Morocco, in 1976. He received the Master's degree in photonics, images, and system control from the Louis Pasteur University, Strasbourg, France, in 2000 and the Ph.D. degree in computer science from the University of Rennes, Rennes, France, in March 2004.

Since 2008, he has been a Researcher with the Computer Vision Laboratory (for more details, see http://labvis.isr.uc.pt/new/index.html), Institute for Systems and Robotics, Coimbra, Portugal (for more details, see http://www.isr.uc.pt/home.php). His current research interests include robotics and computer vision, especially visual servoing.

Youcef Mezouar received the Ph.D. degree in computer science from the Université de Rennes 1, Rennes, France, in 2001 and the "Habilitation à Diriger les Recherches" degree from the Université Blaise Pascal, Clermont-Ferrand, France, in 2009.

For one year, he was a Postdoctoral Associate with the Robotics Laboratory, Computer Science Department, Columbia University, New York, NY. In 2002, he joined the Robotics and Vision Group, LASMEA-CNRS, Blaise Pascal University, where he is currently a Co-lead of the GRAVIR Group (for more details, see http://wwwlasmea.univ-bpclermont.fr) and the ROSACE Team (for more details, see http://wwwlasmea.univ-bpclermont.fr/rosace). His research interests include automatics, robotics, and computer vision, especially visual servoing and mobile-robot navigation.

François Chaumette (SM'10) graduated from École Nationale Supérieure de Mécanique, Nantes, France, in 1987. He received the Ph.D. degree in computer science from the University of Rennes, Rennes, France, in 1990.

Since 1990, he has been with INRIA Rennes, Bretagne Atlantique, Campus Universitaire de Beaulieu, Rennes, where he is currently the "Directeur de Recherches" and the Head of the Lagadic Group (http://www.irisa.fr/lagadic). He is currently a member of the Editorial Board of the *International Journal of Robotics Research*. His research interests include robotics and computer vision, especially visual servoing and active perception.

Dr. Chaumette received the AFCET/CNRS Prize for the best French thesis in automatic control in 1991. He also received, with Ezio Malis, the 2002 King-Sun Fu Memorial Best IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION Paper Award. From 2001 to 2005, he has been an Associate Editor of the IEEE TRANSACTIONS ON ROBOTICS.

Peter Corke received the Ph.D. degree from the University of Melbourne, Melbourne, Vic., Australia, in 1995.

He is currently a Professor with the School of Engineering Systems, Queensland University of Technology, Brisbane, Qld., Australia. He was a Founding Research Director of the Autonomous Systems Laboratory, Commonwealth Scientific and Industrial Research Organisation (CSIRO) ICT Center. He is a member of the Editorial Board of the *International Journal of Robotics Research* and a Founding Editor of the *Journal of Field Robotics*. His current research interests include machine vision, vision-based robot control, field robotics, and sensor networks.

Dr. Corke is currently the Editor-in-Chief of the IEEE ROBOTICS AND AUTOMATION MAGAZINE.